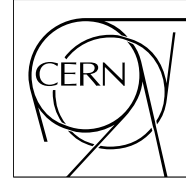


The Compact Muon Solenoid Experiment

**CMS Note**

Mailing address: CMS CERN, CH-1211 GENEVA 23, Switzerland



04 October 2001

# Reconstruction of 20 GeV hard jets in pile-up events at high luminosity at the LHC

O.Kodolova, I.Vardanian

*Moscow State University, Moscow, Russia*

S.Kunori

*University of Maryland, College Park, Maryland, U.S.A.*

A.Nikitenko<sup>1)</sup>

*European Laboratory for Particle Physics (CERN), Switzerland*

## Abstract

The possibility of reconstruction of 20 GeV quark-initiated jets in pile-up background at the high luminosity ( $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ) at the LHC was studied with a modified window jet finder which included an algorithm for event-by-event basis pile-up energy subtraction.  $\sim 74\%$  and  $\sim 70\%$  of jets were reconstructed as a leading jet in the barrel and endcap part of the CMS calorimeter, respectively. The transverse energy threshold value of 0.5 GeV for ECAL tower and HCAL tower is suitable for this jet finder.

---

<sup>1)</sup> On leave from ITEP, Moscow

# 1 Introduction

The reconstruction of jets with low transverse energies plays a crucial role in selection of signal events and rejection of background events in the search for new physics and measurement of their properties at the LHC. For example a typical selection criterion for forward tagging jets associated with Higgs production via weak boson fusion process uses a cut on the jet transverse energy  $E_T$  at 30-40 GeV. Top quark production is copious at the LHC and hence one of the major backgrounds to many new physics channels. A top quark event produces multiple jets in the central pseudorapidity region, while several signal channels do not have jets in the region, such as Higgs productions in which the Higgs subsequently decays into a tau pair or a W pair, and then both W's decaying into leptons and neutrinos. In those cases a central jet veto is very effective to reject the top background, if the central jets can be recognized down to  $E_T$  around 20 GeV.

It is a challenging task to recognize low  $E_T$  jets at the high luminosity ( $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ ) at the LHC, where on average 17.3 minimum bias events are overlapping with a signal event. The event overlap smears the measurement of the jet transverse energy, especially at lower  $E_T$ , and adds extra "jets" due to energy pileup (i.e. fake jets) and real low  $E_T$  jets from the overlapping minimum bias events.

We applied a modified window jet finder to reconstruct light-quark initiated jets with transverse energy  $E_T = 20$  GeV at the high luminosity ( $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ). The original window jet finding algorithm was developed for reconstructing hard jets in heavy ion collisions [1,2,3], where a typical collision dumps transverse energy  $E_T$  up to 10 GeV, depending on ion, in every calorimeter tower. The algorithm was capable of subtracting this background energy during extraction of hard jets on an event-by event basis.

In this note we report the first preliminary result from our studies on jet reconstruction with the modified window algorithm; purity of the leading jets and effect of the  $E_T$  threshold on calorimeter towers. This note deals with only leading jets and more extended studies are in progress.

## 2 Event Generation and Simulation

The signal and pileup minimum bias events were generated with Pythia 5.7 [5] and processed separately through CMSIM 116 [4] to simulate the calorimeter response with GEANT. Then each simulated signal event was merged with minimum bias events to form a pile-up event before applying the modified jet window algorithm for jet reconstruction.

As a signal we considered the hard parton-parton scattering process ( $q + \bar{q} \rightarrow q + \bar{q}$ ), which was generated using subroutines LU2ENT, LUROBO and LUSHOW (Pythia 5.7) [5]. LU2ENT generated partons with an exact value of  $E_T$ , LUROBO was allowed to choose the exact direction of the partons and LUSHOW enabled the initial and final state radiation.

2000 minimum-bias background events at  $\sqrt{s} = 14$  TeV were generated with Pythia 5.7, using "standard" settings [6], which included:

- The parton density structure function CTEQ 4L,
- Pythia generation datacards (Table 1).

**Table 1. Pythia generation datacards for minimum-bias events**

MSEL = 1	Non-diffractive (jet) processes
MSTP(2) = 2	2nd order runnings $\alpha_s$
MSTP(33) = 3	$K$ -factors: fixed and $\alpha_s$ scale shift at $\text{PARP}(33) = 0.075 * Q^2$
MSTP(81) = 1	Multyi-parton interactions switched on
PARP(82) = 3.2	regularization scale $p_{\perp 0}$ of the transverse momentum spectrum for multiple interaction with $\text{MSTP}(82) \geq 2$

The average charged particle multiplicity of the minimum bias events was  $dN^{\pm}/dy$  ( $y = 0$ ) = 7.1 per unit pseudorapidity and the average transverse momentum was  $\langle p_{\perp} \rangle = 0.61 \pm 0.58$  GeV for charged particles with  $p_{\perp} > 150$  MeV in the region of pseudorapidity  $|\eta| < 2.5$ .

CMS calorimeter responses were simulated using detailed calorimeter geometry description (CMSIM116

[4]). The barrel and endcap part of the calorimeter were considered. We used the following segmentation of calorimeter towers, which includes both electromagnetic (ECAL) and hadronic (HCAL) parts. In all of the barrel and most of the endcap regions, the size of the tower was  $\Delta\eta = 0.0870$  by  $\Delta\varphi = 2\pi/72 \approx 0.0873$ . At high  $\eta$  in the endcap region ( $\eta > 1.74$ ), the towers became larger in  $\eta$ . In the following text we use the term "cell" which coincides with calorimeter tower.

The signal event sample and minimum-bias event sample were processed separately event by event through CMSIM to produce HITS. GHEISHA was used to simulate hadronic shower in the detector.

Mixing HITS of signal event with HITS of minimum-bias background events to form pile-up events was performed at the level of digitization. Minimum bias events,  $\langle N_{mb} \rangle = 17.3$  per bunch crossing, were added to a signal event according to Poisson distribution. The number of the pile-up events was 1000.

After the event mixing, electronics noise was added to each readout channel. The noise levels were 0.03 GeV ( $E_T$ /crystal) and 0.150 GeV ( $E$ /crystal) for the ECAL barrel and ECAL endcap, respectively. The added noise for HCAL was negligible. Thresholds cuts were applied to each ECAL readout channel; .03 GeV ( $E_T$ /crystal) and 0.150 GeV ( $E$ /crystal) for the barrel and endcap, respectively.

The HCAL calibration constants (shown in table 2) were used to translate energy deposited in the scintillator in HCAL to estimated energy in HCAL. These constants for GHEISHA were determined with  $E_T = 50$  GeV pions at  $0.05 < \eta < 0.3$  for the barrel and  $1.8 < \eta < 2.2$  for the endcap part of the calorimeter[7]. The GEANT CUTS were 1 MeV for electrons and gammas and 10 MeV for hadrons.

**Table 2. Calibration constants for hadron calorimeter**

Barrel:	72E5	147E5	147E5	156E5
Endcap:	108E5	237E5	237E5	

### 3 Window jet finding algorithm

The modified window-type jet finding algorithm was used to search "jet-like" clusters above the average energy.

- First the average transverse energy  $\overline{E_T^{cell}(\eta)}$  as a function of  $\eta$  and dispersion  $D_T^{cell}(\eta)$  were calculated over all cells in the barrel and endcap calorimeters as a function of pseudorapidity  $\eta$ , where the dispersion was defined as  $D_T^{cell}(\eta) = \sqrt{(\overline{E_T^{cell}(\eta)})^2 - \overline{(E_T^{cell}(\eta))^2}}$

- Then all possible rectangular windows with a fixed number of cells were created by sliding a window by one cell through an  $\eta - \varphi$  grid of calorimeter cells. The numbers of cells in windows in  $\eta$  ( $N_\eta^{wind}$ ) and in  $\varphi$  ( $N_\varphi^{wind}$ ) were calculated separately by the expressions:

$$\begin{aligned} N_\eta^{wind} &= R * N_\eta^{total} / \eta_{max}, \\ N_\varphi^{wind} &= R * N_\varphi^{total} / 2\pi, \end{aligned}$$

where  $N_\eta^{total}$  and  $N_\varphi^{total}$  were the total numbers of cells in the calorimeters in  $\eta$  and  $\varphi$ ,  $\eta_{max}$  is the maximal value of pseudorapidity  $\eta$ , and  $R$  is an external parameter of the algorithm.

- The window energy was calculated as the sum of the cell transverse energy  $E_T^{cell}$  for all cells ( $n_c$ ) in the window minus the background energy per cell:

$$E_T^{wind} = \sum_{n_c} \{E_T^{cell} - [\overline{E_T^{cell}(\eta)} + D_T^{cell}(\eta)]\}$$

If the value of the transverse cell energy after subtraction of the background energy per cell became negative it was set to zero. This energy subtraction scheme and treatment of negative energy were found to be optimal for heavy ion case [1,2,3] and used in this analysis without any additional tuning.

Then the search for jets and evaluation of the jet energies were started from a window with the maximum transverse energy.

- The non-overlapping windows with energy  $E_T^{wind} > 2\sqrt{\sum D_T^{cell}(\eta)^2}$  were considered as candidates for jets, where the summation was done over all cells in the window.

- The center of gravity of the window was considered as the center of the jet.
- For correction of the jet axis a cell with maximum transverse energy in cone is found and considered as a new geometrical center of this jet. Cells within radius  $R$  around the new geometrical center are collected and center of gravity of jet is recalculated. This correction procedure was optional, and could be applied or excluded from the algorithm.
- Cells in a cone with radius  $R$  around jet center were collected.
- The value  $\overline{E_T^{cell}(\eta)}$  and  $D_T^{cell}(\eta)$  were recalculated using cells which were not covered by jets.
- The jet energy was calculated as energies in the collected cells minus mean background energy per cell:

$$E_T^{jet} = \sum \{E_T^{cell} - [\overline{E_T^{cell}(\eta)} + D_T^{cell}(\eta)]\}.$$

## 4 Transverse energy flow in background pileup events

The  $E_T$  flow was studied for pure background pile-up events, i.e. no signal event was included in the calculation. The energy flow was defined as the density of transverse energy per unit pseudorapidity ( $\langle \Delta E_T^{cell}(\eta) / \Delta \eta^{cell} \rangle$ ), where  $\Delta E_T^{cell}(\eta)$  was the transverse energy averaged over all cells around the whole azimuth at a given  $\eta$ , and  $\Delta \eta^{cell}$  was the  $\eta$  size of the cell. The  $E_T$  density as a function of  $\eta$  is shown in figure 1.

The transverse energy flow showed clear  $\eta$  dependence in the endcap region ( $\eta > 1.5$ ), but not in the barrel. The averaged values of  $E_T$  per cell were: 0.06 GeV and 0.02 GeV for ECAL and HCAL in the barrel, and 0.13 GeV and 0.09 GeV for ECAL and HCAL in the endcap, respectively.

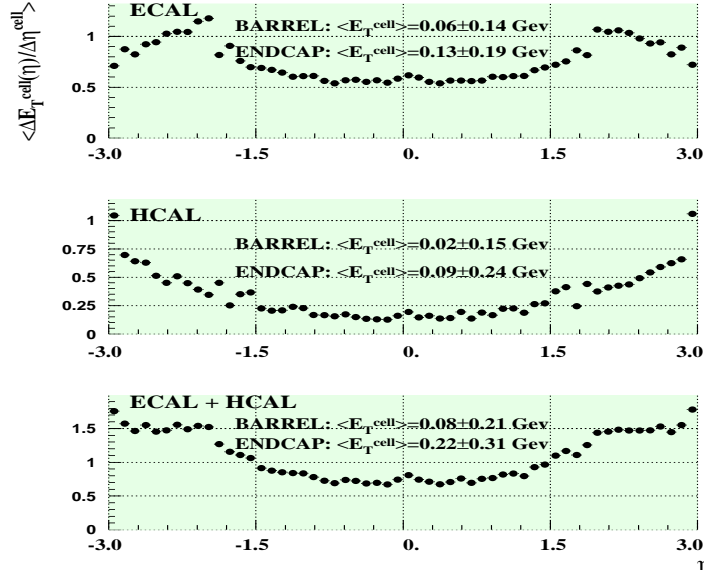


Figure 1: Dependence on pseudorapidity  $\eta$  of the transverse energy density averaged over  $\varphi$  for minimum bias background events.

## 5 Jet reconstruction

Jet reconstruction was studied using the window algorithm under the following conditions:

- hard events  $q + \bar{q} \rightarrow q + \bar{q}$ ;
- initial energy of parton  $E_T = 20$  GeV;
- with and without background from pile-up (high luminosity);

- barrel and endcap calorimeters considered;
- with and without thresholds on ECAL and HCAL cell energy;
- cut on window energy:  $E_T^{wind} > 5$  GeV;
- only one reconstructed jet with the largest transverse energy was used for further analysis.

At first, we considered jets reconstructed in the pure pile-up events (one pile-up event was taken as  $\langle 17.3 \rangle$  minimum-bias events) and within pseudorapidity region  $|\eta| < 3.0$ . We call such jets as “false” jets. The distribution of  $E_T$  of the leading “false” jet reconstructed without and with background subtraction (cone size 0.6) is shown in figures 2 a and 2 b. One can see that background subtraction decreased the average energy of the “false” jets by a factor of 2 - from 30 GeV to 15 GeV.

A sample of “false” jets consisted of both fake jets due to overlap of soft particles from different minimum-bias events and real hard jets from minimum-bias events. At the level of the event generator 6 % of minimum bias events had at least one hard jet with  $E_T > 16$  GeV. For a pure pile-up event sample of  $\langle 17.3 \rangle$  minimum-bias events, we estimated that 66 % of the pile-up events would have at least one hard jet with  $E_T > 16$  GeV. The estimation was made using the following formula:

$$P(E_T \geq 16 \text{ GeV}) = 1 - (1 - p)^{17.3}, \quad (1)$$

where  $p (= 0.06)$  is the probability to have at least one jet with  $E_T \geq 16$  GeV in a minimum-bias event.

In a simulated event sample of 300 minimum-bias events (single interaction) we found the following numbers of reconstructed “false” jets (table 3).

- **Without background subtraction.** The fraction of jets with  $E_T \geq 16$  GeV was 4 %. The distribution of multiplicity of jets with  $E_T \geq 16$  GeV is given in table 3. Using equation 1, we obtained  $P(E_T \geq 16 \text{ GeV}) = 0.5$  or 50%.
- **With background subtraction.** The fraction of jets with  $E_T \geq 16$  GeV was 2 %. The distribution of multiplicity of jets with  $E_T \geq 16$  GeV is given in table 3. Using equation 1, we obtained  $P(E_T \geq 16 \text{ GeV}) = 0.29$  or 29%.

**Table 3. Multiplicity of jets with reconstructed transverse jet energy  $\geq 16$  GeV by the the window algorithm with  $R = 0.6$  without and with background subtraction for minimum-bias events.**

multiplicity	without subtraction	with subtraction
0	289	295
1	6	4
2	3	1
3	1	0
4	1	0

In a pure pile-up event sample ( $\langle 17.3 \rangle$  minimum-bias events), we found 97 % ( $34 \pm 6\%$ ) of events had jets with  $E_T \geq 16$  GeV. The mean multiplicity of such jets was 8 (0.9) without (with) background subtraction. This calculations was made with 100 pile-up events created from a sample of 2000 minimum-bias events. Each minimum bias event was only used once to form pile-up events. Comparing these fractions of events with  $E_T > 16$  GeV and the estimated probability  $P(E_T \geq 16 \text{ GeV}) = 50$  % (29 %), we conclude that half of the events with jets with  $E_T > 16$  GeV without background subtraction were due to real hard jets, while more than 80 % of them with the background subtraction was due to real hard jets.

For hard ( $q + \bar{q} \rightarrow q + \bar{q}$ ) events with and without pile-up background, Table 4 shows the mean transverse energy for the reconstructed leading jets with cone size 0.6 with background subtraction and no threshold for ECAL and HCAL cell energy. While the background subtraction reduced the mean transverse energy of leading jets by a factor of 2 in the case of the pure pile-up event sample, it kept the mean transverse energy of leading jets at about the same value after adding pile-up events in the case of the signal event sample.

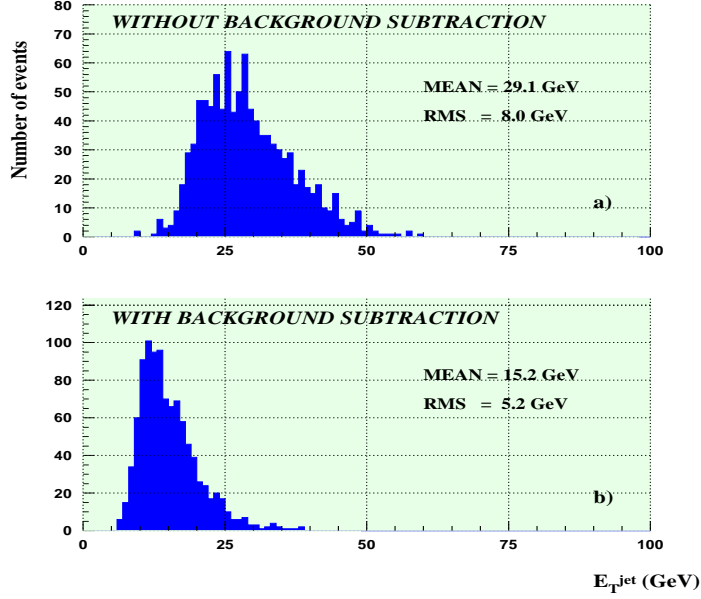


Figure 2: The transverse energy distribution of the "false" jets extracted from pile-up only ( $\langle 17.3 \rangle$  min. bias) in the region  $|\eta| \leq 3.0$  without (a) and with (b) background subtraction.

Table 4. The mean transverse energy of the reconstructed leading jets with the window algorithm with  $R = 0.6$  for the signal samples without and with pile-up background.

Reconstructed $E_T^{\text{jet}}$ (GeV)		
Background	BARREL	ENDCAP
without pile-up	$13.6 \pm 4.2$	$17.2 \pm 4.2$
with pile-up	$15.5 \pm 5.2$	$18.1 \pm 5.4$

## 6 Purity of leading jets

The window algorithm always found at least one jet with  $E_T > 5$  GeV in the signal event sample with pile-up. The leading jets would consist of both "true" from signal and "false" jets from overlapping minimum bias events.

In order to identify the "true" jets, we reconstructed jets from the events without pile-up background and marked the cells in the jet cone. Then we looked for the same jet in the event with pile-up background and determined the number of cells overlapping with previously marked cells. Figure 3 and Table 5 shows the fraction of events as a function of the percentage of overlapping cells for jets reconstructed by the window algorithm with  $R = 0.6$ .

89% and 80% of the leading jets have at least one overlapping cell in the barrel and endcap, respectively. The low amount of overlapping cells meant that jet's direction was shifted. So further we defined "true" jets as only those jets with greater than 60% of cells overlapping. The fractions of "true" jets were 74% and  $\sim 70\%$  for barrel and endcap, respectively (Table 5). 11% of events in the barrel and 20% in the endcap did not have overlapping cells at all. These were really "false" jets.

## 7 Thresholds for ECAL and HCAL cell energy

One of the possible scenarios of the calorimeter readout is to set threshold values in the HCAL and ECAL readout electronics and read only the cells with  $E_T^{\text{cell}}$  above the threshold values and send them to the DAQ to record the energy in those cells on tape. At most 15% of total cells are assumed to be recorded on tape.

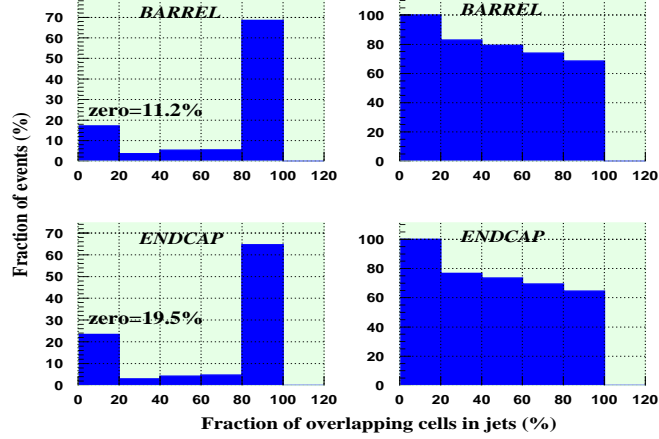


Figure 3: The fraction of events vs. the fraction of overlapping cells. Jets were reconstructed using the window algorithm with  $R = 0.6$ . Differential (left) and integral (right) distributions. ZERO indicates the fraction of events without any overlapping cell.

**Table 5. The fraction of events with given percentage of overlapping cells**

$R_{algo} = 0.6$ , without threshold for cell energy	
Overlapping cells (%)	Events (%)
BARREL	
> 60	74
< 60	15
0	11
ENDCAP	
> 60	69
< 60	11
0	20

We studied the influence of the threshold values on the performance of the window jet finding algorithm. Three  $E_T$  thresholds, 0.0, 0.5 and 1.0 GeV were applied separately to ECAL tower and HCAL tower. The definition of the ECAL tower was taken from the L1 trigger tower definition; combined 5x5 crystals formed a tower in the barrel and appropriate crystals matching to HCAL tower in the endcap.

Shown in figure 4 are the fraction of events with "true" and "false" jets reconstructed for high luminosity using the window algorithm with different radii  $R$  ( $R = 0.6, 0.7, 1.0$ ) as a function of the threshold values.

With the threshold values increasing from 0.5 GeV to 1 GeV, the fraction of events with "true" jets decreased from  $\sim 75\%$  to  $\sim 70\%$  for the barrel and from  $\sim 70\%$  to  $62\%$  for the endcap part of calorimeters, and did not depend on the jet cone size  $R$ . On the contrary, the contribution of events with "false" jets increased with increasing threshold. The "false" jet rates strongly depended on the cone size  $R$ . The bigger the jet cone size, the smaller the fraction of events with "false" jets. Also the "false" jet rate is expected to be depend on a  $E_T$  cut of jets. More detailed study on characteristic of "false" jets and algorithm to reduce the "false" jet rate is in progress.

The effect of the threshold cuts on jet  $E_T$  resolution ( $\sigma((E_T^{jet} - E_T^{parton})/E_T^{jet})$ ) is presented in figure 5. The best energy resolution for "true" jets will be obtained without any threshold cuts. For "true" jets reconstructed with threshold value 0.5 GeV and the jet cone size  $R = 0.6$  the resolution in jet transverse

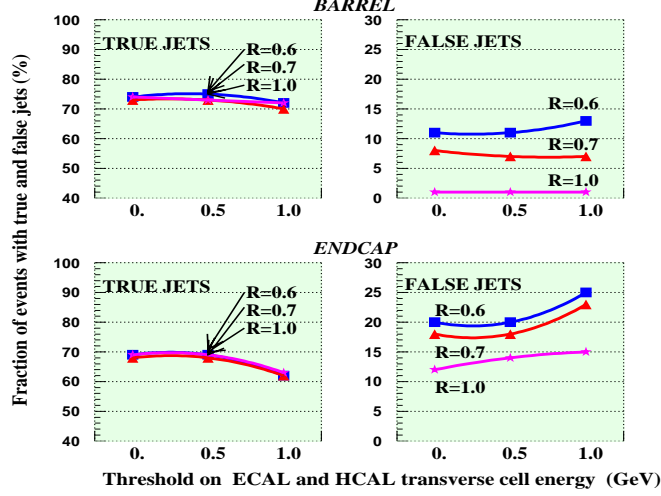


Figure 4: The fraction of events with "true" (left) and "false" (right) jets reconstructed using the window algorithm with different radii  $R$  ( $R = 0.6, 0.7, 1.0$ ) in the barrel (upper) and the endcap (lower) as a function of threshold values in  $E_T$  for ECAL and HCAL towers.

energy was  $\sim 38\%$  for the barrel and  $30\%$  for the endcap. The resolution became worse with increasing threshold, but improved with the jet cone size,  $R$ .

The dependence of transverse energy smearing on threshold value is shown in figure 6. The smearing is the relative differences in transverse energy between the "true" jets reconstructed in events with pile-up and threshold and jets reconstructed in events without pile-up and threshold. One can see in figure 6 that the best reconstruction was with threshold of tower  $E_T = 0.5$  GeV. With the radius of the algorithm,  $R = 0.6$ , and the threshold value  $0.5$  GeV, the smearing was  $\sim 5\%$  for the barrel and  $\sim 3\%$  for the endcap.



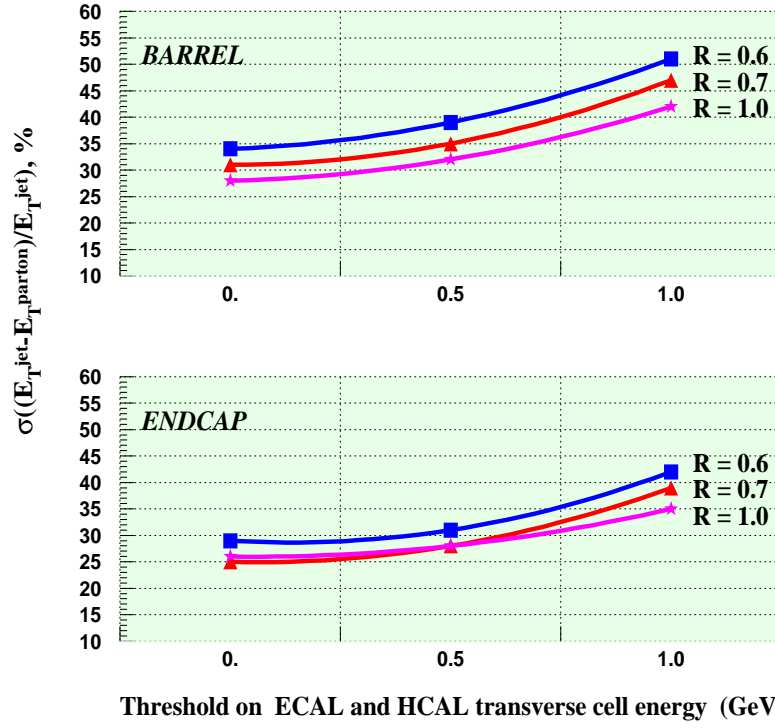


Figure 5: The energy resolution ( $\sigma((E_T^{jet} - E_T^{parton})/E_T^{jet})$ ) of "true" jets reconstructed using window algorithm with different radii ( $R = 0.6, 0.7, 1.0$ ) in barrel (upper) and endcap (lower) region of calorimeters as a function of threshold value for ECAL and HCAL transverse cell energy.

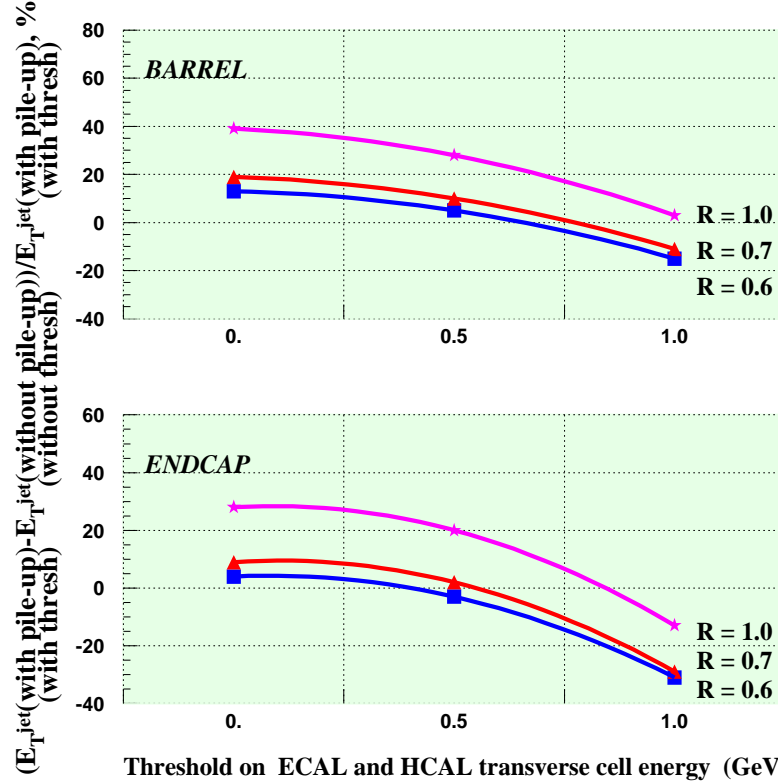


Figure 6: The relative differences in transverse energy between the "true" jets reconstructed in events with pile-up and threshold and jets reconstructed in events without pile-up and threshold in barrel (upper) and endcap (lower) of calorimeters as a function of threshold value for ECAL and HCAL transverse cell energy.

According to the presented results, the optimum threshold values for ECAL and HCAL towers are around

0.5 GeV in  $E_T$ .

## 8 Conclusions

The possibility of jet reconstruction with the window jet finding algorithm for hard  $q + \bar{q} \rightarrow q + \bar{q}$  events with initial transverse energy of the parton  $E_T = 20$  GeV in the pile-up background at high luminosity ( $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ) was investigated.

At least one jet with  $E_T > 5$  GeV was found in every event. 89% and 80% of the leading jets reconstructed with the jet cone size  $R = 0.6$  and the threshold cuts at  $E_T = 0.5$  GeV on the ECAL tower and HCAL towers had at least one cell overlap with the same jets in events without pile-up background in the barrel and endcap part of the calorimeter, respectively.

If we look for "true" reconstructed jets in events with pile-up background, defined as jets with more than 60% of overlapping cell with the same jets in events without pile-up background, the corresponding purity became 74% and 70%.

Setting the threshold value to 0.5 GeV for each tower of ECAL and HCAL calorimeters seems the most suitable for jet reconstruction.

This note described results on only leading jets. More extended study is in progress to understand the influence of the pile-up energy subtraction for all reconstructed jets on: a) resolution of true jets, b) energy scale of true jets, c) reduction of false jet rate. Also, most of the parameters in the window jet finding algorithm were inherited from those for heavy ion environment. We plan to do further optimization for various luminosity at the LHC.

## 9 Acknowledgments

We are thankful to L.Sarycheva, N.Kruglov, A.Demianov and A.Gribushin for useful discussions, and V.Gavrilov for the interest and support.

## References

- [1] N.Kruglov et al. "*Jet finding algorithm in heavy ion collisions in CMS*" CERN CMS TN/96-084.
- [2] M.Bedjidian et al. "*Jet Physics in CMS Heavy Ion Program*" CERN CMS NOTE 1999/016.
- [3] A.Nikitenko et al. "*Study of Gamma+Jet Channel in Heavy Ion Collision with CMS*" CERN CMS NOTE 1998/063.
- [4] CMSIM user's Guide at [www.http://cmsdoc.cern.ch/swsi.html](http://cmsdoc.cern.ch/swsi.html).
- [5] T.Sjostrand. *Comp. Phys. Com.* 82/1994, 74.
- [6] C.Seetz. "*Minimum-Bias Pileup Issues in Electron-Photon HLT studies*" CERN CMS IN 2000/001.
- [7] S.Kunori. [http://home.fnal.gov/~kunori/cms\\_sw/colo\\_sw/hcal\\_hits\\_digi.html](http://home.fnal.gov/~kunori/cms_sw/colo_sw/hcal_hits_digi.html).